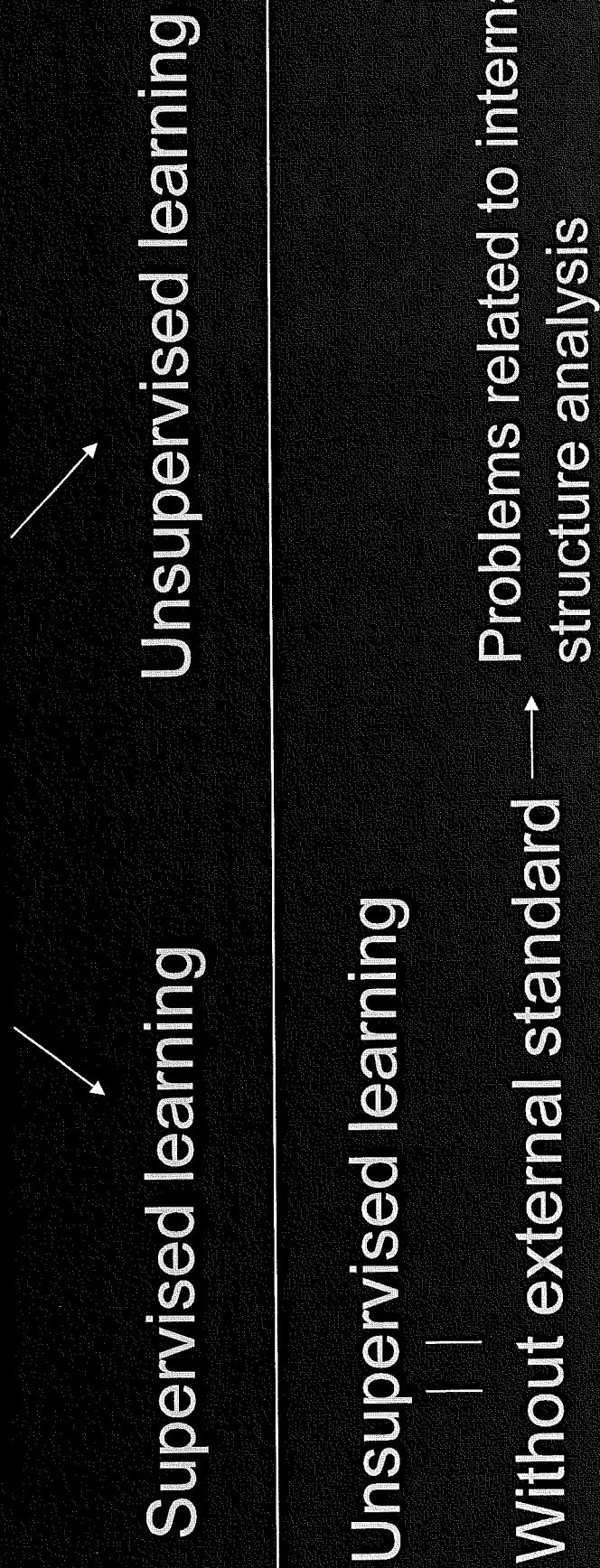


Methods of Summarizing Data



Common in biotechnology

Cluster analysis : Data are summarized by gathering similar data.
Principal component analysis : Data are summarized such that the loss of original information is minimized.

Psychology

Factor analysis : Data are summarized based on a mathematical model.

Need for Supervised Learning

Supervised learning



With external standard →

Problems related to prediction,
discrimination, or diagnosis

Comparison of two groups is needed for realization of
personalized medicine.

< Examples of comparison >

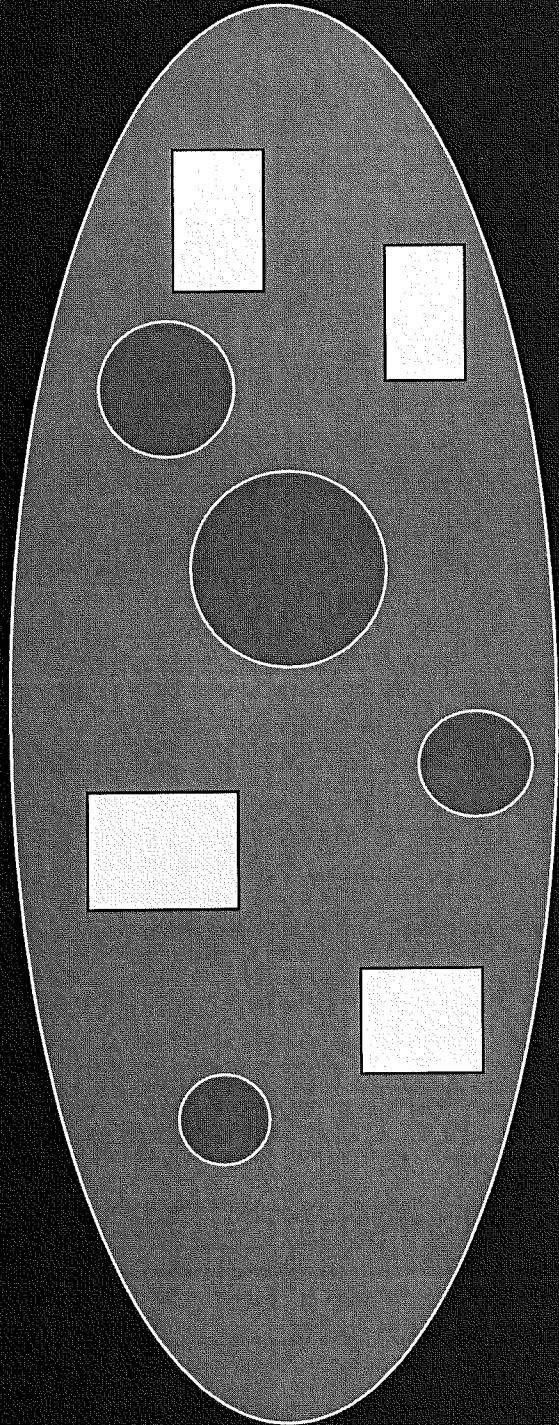
1. Normal tissue vs. Cancer tissue
2. Primary cancer vs. Metastatic cancer
3. Before agent administration vs. After agent administration
4. Pre-radiation vs. Post-radiation
5. Recurrent cancer vs. Non-recurrent cancer

Supervised Learning vs. Unsupervised Learning

Unsupervised learning (conventional technique)

Cancerous samples ●

Non-cancerous samples ■

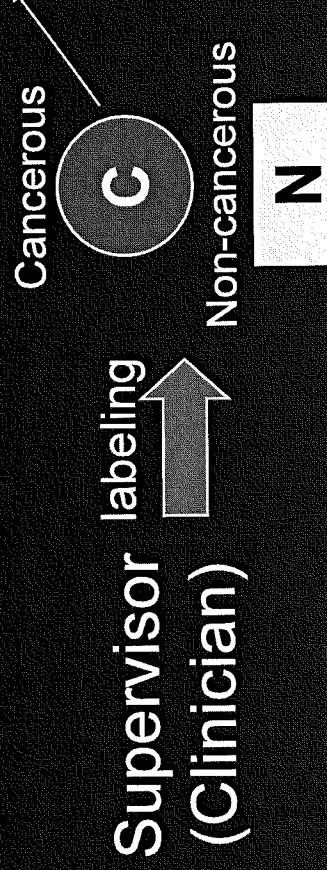


Analysis is performed on the mixture of cancerous and non-cancerous samples.

Supervised Learning vs. Unsupervised Learning

Supervised learning

Using samples that are labeled
by a supervisor in advance



Analysis is performed by comparing cancerous and non-cancerous samples.